1 – 23 (Cancelled).

23. (new) A local videoconferencing device for a videoconferencing system having a local videoconferencing device with a video display and at least one remote videoconferencing device with a video display interconnected through a network, the local videoconferencing device comprising:

a video sensor for capturing images;

a plurality of microphones for capturing sound, the plurality of microphones being arranged in known positions relative to one another;

a plurality of speakers for producing sound, the plurality of speakers being arranged in known positions relative to one another;

at least one processing unit coupled to the video sensor, the microphones and the speakers; and

a communication interface coupled to the at least one processing unit and the at least one remote videoconferencing device through the network;

wherein the at least one processing unit is operative to produce at least a first video stream from signals received from the video sensor and an audio stream and an audio source position signal from signals received from the microphones, wherein the audio source position signal is based upon the magnitude differences of captured sound from the plurality of microphones;

wherein the at least one processing unit is operative to receive at least one video stream, one audio stream, and one audio source position signal from a remote videoconferencing device; and

wherein the at least one processing unit is operative to drive the plurality of speakers to reproduce sound according to the received audio stream and audio source position signal.

24. (new) The videoconferencing device of claim 23, wherein the video sensor is operative to produce high resolution video stream, wherein the first video stream is of a first resolution, wherein the at least one processing unit is operative to produce a second video stream, and wherein the second video stream is of a second resolution and is representing an area in the first video stream.

25. (new) The videoconferencing device of claim 24, wherein the first resolution of the first video stream is 700x400 pixels, and wherein the second resolution of the second video stream is 300x200 pixels.

26. (new) The videoconferencing device of claim 24, wherein the maximum resolution of the video sensor is 3000x2000 pixels.

27. (new) The videoconferencing device of claim 24, wherein the second video stream represents images of a speaking videoconference participant.

28. (new) The videoconferencing device of claim 27, wherein the second video stream follows the speaking videoconference participant and changes when the speaking videoconference participant changes.

29. (new) The videoconferencing device of claim 23, wherein the at least one processing unit is operative to synchronize the phases of the signals from the video sensor and a video stream output by a remote videoconference device for display on a remote video display.

30. (new) The videoconferencing device of claim 23, wherein the at least one processing unit is operative to drive the plurality of speakers to reproduce sound according to the received audio signal and audio source position signal by selectively driving one or more speakers in response to the received position signal from the remote videoconferencing device to play the audio signal corresponding to the image of the at least one video stream.

31. (new) The videoconferencing device of claim 23, wherein the video sensor has a wide viewing angle.

32. (new) The videoconferencing device of claim 31, wherein the wide viewing angle is 65 degrees.

33. (new) The videoconferencing device of claim 31, further comprising a pan motor to increase the viewing angle of the video sensor.

34. (new) A method for videoconferencing, wherein a plurality of videoconferencing devices are interconnected through a network, wherein each videoconferencing device comprises a video sensor, a plurality of microphones and speakers, a processing unit, a video display and a network interface, the method comprising:

    capturing video images with the video sensor;

    capturing audio signals with the microphones;

    receiving the video images and the audio signals at the processing unit;

    generating a first video stream from the video images and an audio stream and an audio position signal from the audio signals, wherein the audio position signal is generated based upon magnitude differences of audio signals received from the plurality of microphones;

    transmitting the first video stream, audio stream and audio position signal to a remote conferencing device.

35. (new) The method in claim 34, wherein the video images are of high resolution, wherein the first video stream is of a first resolution.

36. (new) The method in claim 35, further comprising the processing unit generating a second video stream, wherein the second video stream is of a second resolution and is representing an area in the first video stream.

37. (new) The method in claim 36, wherein the second video stream represents images of a speaking videoconference participant.

38. (new) The method in claim 34, wherein the processing unit synchronizes phases of the signals.

39. (new) The method in claim 34, further comprising:

    receiving at least one remote video stream, at least one remote audio stream, and a remote audio position signal from a remote endpoint;

    displaying the at least one remote video stream on the video display; and

    driving the plurality of speakers to reproduce sound by selectively driving one or more of the plurality of speakers in response to the received remote audio position signal.

40. (new) A method for videoconferencing, wherein a plurality of videoconferencing devices are interconnected through a network, wherein each videoconferencing device comprises a video sensor, a plurality of microphones and speakers, a processing unit, a video display and a network interface, the method comprising:

    receiving at a local endpoint via the network interface at least one remote video stream, at least one remote audio stream, and a remote audio position signal from a remote endpoint;

    displaying the at least one remote video stream on the video display; and

    driving the plurality of speakers to reproduce sound by selectively driving one or more of the plurality of speakers in response to the received remote audio position signal, wherein the received remote audio position signal is generated based upon magnitude differences of audio signals received at a plurality of remote microphones.